R E P O R T R E S U M E S

ED 019 410 VT 001 835

DEVEOPMENT OF A PERFORMANCE SPECIMEN CHECK LIST FOR MACHINING PERSONNEL.

BY- BOSTOCK, D.J.
UNION CARBIDE CORP., OAK RIDGE, TENN.,NUCLEAR DIV.

REPORT NUMBER V-1461 PUB DATE 20 APR 64
REPORT NUMBER TID-4500
DOCUMENT NOT AVAILABLE FROM EDRS. EDRS Price MF $0.50 — 57P.

DESCRIPTORS- METHODS RESEARCH, *PERSONNEL EVALUATION, *MACHINE TOOL OPERATORS, *CRITICAL INCIDENTS METHOD, *PERFORMANCE FACTORS, *RATING SCALES, EVALUATION TECHNIQUES, OBSERVATION,

A MODIFIED CRITICAL INCIDENT PERFORMANCE EVALUATION FORM FOR HOURLY MACHINING PERSONNEL WAS DEVELOPED, AND CERTAIN RATER AND RATING INSTRUMENT CHARACTERISTICS WERE INVESTIGATED FROM THE STANDPOINT OF THEIR EFFECTS ON THE VARIANCE OF PERFORMANCE. INTERVIEWS WERE CONDUCTED WITH FRONTLINE FOREMEN TO COLLECT PERFORMANCE SPECIMENS WHICH WERE EVALUATED AND INCORPORATED INTO A RATING FORM AND A RATING SCALE FOR OVERALL PERFORMANCE. THE FORM, CONSISTING OF 26 EFFECTIVE AND 26 INEFFECTIVE SPECIMENS, WAS TESTED BY 22 FOREMEN RATING A SAMPLE GROUP OF 198 MACHINISTS. THE RATIO OF EFFECTIVE TO INEFFECTIVE SPECMENS PROVED TO BE RELIABLE AND A RELEVANT MEASURE OF OVERALL PERFORMANCE. THE REGRESSION EQUATION OF THE EFFECTIVE-INEFFECTIVE VERSUS RATING SCORE WAS CONSISTENT WITH PREVIOUS RESULTS. A CRITICALITY FACTOR WAS OBTAINED FOR EACH SPECIMEN FROM THE AVERAGE FOREMEN WEIGHTS. THE INCORPORATION OF THE CRITICALITY FACTORS (THE MEASURES OF THE IMPORTANCE OF A PERFORMANCE SPECIMEN IN RELATION TO THE TOTAL SPECIMEN POPULATION) IN CALCULATING THE SPECIMEN RATIO DID NOT REDUCE THE VARIANCE OF ESTIMATE IN PREDICTING EVALUATION FROM OBSERVATION. BY REDUCING THE SPECIMEN LIST TO ITEMS WITH HIGH CRITICALITY FACTORS, THE EFFECTIVE-INEFFECTIVE RATIO WAS SIGNIFICANTLY REDUCED. THE INCREASE IN THE VARIANCE OF ESTIMATE DUE TO RATER DIFFERENCES DRAMATIZES THE NEED FOR A GENERAL EQUATION THAT WILL ESTIMATE AVERAGE PERFORMANCE EVALUATION RESPONSE FOR A GIVEN SET OF OBSERVED PERFORMANCES. THE FORM DEVELOPED IS RELIABLE AND RELEVANT AND MAY BE USED IN THE INDUSTRIAL COMPLEX FOR WHICH IT WAS DESIGNED. THE FORM IS INCLUDED. THIS DOCUMENT IS AVAILABLE FOR $1.50 FROM CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION, U.S. DEPARTMENT OF COMMERCE, 5285 PORT ROYAL ROAD, SPRINGFIELD, VIRGINIA 22151. (HC)

Y-12 CENTRAL FILES

Y-1461
General, Miscellaneous,
and Progress Reports

cy. 64
67

AEC RESEARCH AND DEVELOPMENT REPORT

ED0194410

DEVELOPMENT OF A PERFORMANCE SPECIMEN CHECK LIST FOR MACHINING PERSONNEL

D. J. Bostock

Y-12 PLANT
Oak Ridge, Tennessee

UNION CARBIDE

UNION CARBIDE CORPORATION
NUCLEAR DIVISION

Operating the

OAK RIDGE GASEOUS DIFFUSION PLANT • OAK RIDGE NATIONAL LABORATORY
OAK RIDGE Y-12 PLANT • PADUCAH GASEOUS DIFFUSION PLANT

For the Atomic Energy Commission
Under U.S. Government Contract W7405 eng 26

VT001835

Printed in USA. Price: $1.50 Available from the
Office of Technical Services
U. S. Department of Commerce
Washington 25, D. C.

LEGAL NOTICE

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Date Issued: April 20, 1964

Report Number Y-1461

General, Miscellaneous,
and Progress Reports
TID-4500 (27th Edition)

UNION CARBIDE CORPORATION
Nuclear Division

Y-12 PLANT

Contract W-7405-eng-26
With the US Atomic Energy Commission

DEVELOPMENT OF A PERFORMANCE SPECIMEN CHECK LIST FOR MACHINING PERSONNEL.

D. J. Bostock

This report is based on a study by the author as partial fulfillment of requirements for the degree of Master of Science in Industrial Management from The University of Tennessee.

Oak Ridge, Tennessee
February 19, 1964

Report Number Y-1461

General, Miscellaneous,
and Progress Reports
TID-4500 (27th Edition)

Distribution:

Bailey, E. W.
Ballenger, H. F. (2)
Bell, B. B.
Bernander, N. K.
Bostock, D. J. (5)
Burkhart, L. E.
Cable, R. E.
Center, C. E. (ORGDP)
Christman, A. M.
Cowen, D. D. (ORNL)
Dow, N.
Evans, G. W.
Fortenbery, M. J.
Googin, J. M.
Gritzner, V. B.
Harding, J. E.
Harwell, W. L. (ORGDP)(5)
Hemphill, L. F.
Hensley, C. E.
Huber, A. P. (ORGDP)
Keller, C. A. (AEC-ORO)(4)
Kite, H. T.
Lewis, A. D.
Lewis, F. O. (ORGDP)
Little, J. C.
Lockhart, G. B.
McLendon, J. D.
Mitchel, G. W.
Patton, F. S.
Perry, A. E.
Rader, D. H.
Schwenn, M. F.
Sommerfeld, K. W.
Snyder, H. G. P. (3)
Stoner, H. H.
Tench, F. M.
Thomas, W. F.
Tilson, F. V.
Trotter, T. C.
Waters, J. L.
Webber, T. R.
Whitson, W. K.
Wilkinson, P. E.
Williams, J. L.
Winkel, R. A. (Paducah)
Zurcher, E.
Y-12 Central Files (5)
Y-12 Central Files (Y-12RC)

In addition, this report is distributed in accordance with the category, General, Miscellaneous, and Progress Reports as given in the "USAEC Standard Distribution Lists for Unclassified Scientific and Technical Reports", TID-4500 (27th Edition), February 1, 1964.

ABSTRACT

This study was the development of a modified critical incident performance evaluation form for hourly machining personnel and an investigation of factors which contribute to the variance of estimate of evaluation of performance from observation of performance.

Interviews were conducted with frontline foremen to collect performance specimens which were then incorporated into a rating form along with a rating scale for overall performance. The form was tested by 22 foremen rating a sample group of 198 machinists. The ratio of effective to ineffective specimens proved to be reliable and a relevant measure of over-all performance. The independent rater reliability was .663 while the correlation of the ratio and a rating scale value was .700. The regression equation of the E/I vs rating score was consistent with previous results although the slope was less due to a small range of the criterion.

Weights of the importance of the performance specimens as assigned by foremen were found to be low in reliability but the average assignment correlated with Uhrbrock's scale values. A criticality factor was obtained for each specimen by predicting an Uhrbrock value from the average foremen weights.

The incorporation of the criticality factors in calculating the specimen ratio did not reduce the variance of estimate in predicting evaluation from observation.

By reducing the specimen list to items with high criticality factors the E/I ratio was significantly reduced. This is attributed to a reduction of redundancy and the "halo" effect operating between a portion of the effective specimens.

Rater differences added significantly to the variance of estimate; however, they could not be explained by attitude or ability differences as measured by the foremen's supervision.

ACKNOWLEDGEMENTS

I am deeply indebted to the many who provided data and assistance in the completion of this thesis. A particular thanks is extended to the supervision of the Fabrication Division of the Y-12 Plant and to Professor G. H. Whitlock for his direction of the entire work.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	General Introduction	1
	Statement of the Problem	4
	Importance of the Problem	5
	Definition of Terms	5
	Approach to the Problem	6
	Organization of Thesis	7
II.	HISTORY OF THE PROBLEM	8
	Critical Incident Technique and the Psychophysical Law	8
	Criticality Factors	10
III.	METHODOLOGY	12
	Develop Check List and Form	13
	Criticality Factors	13
	Sample Selection	14
	Rating the Sample	15
	Scoring	15
	Techniques of Analysis	16
IV.	RESULTS	19
	The Form	19
	Criticality Factors	19

CHAPTER PAGE

Scoring 21

Unweighted Data 25

Attenuated Data 28

Weighted Data 28

Rater Effect 31

V. DISCUSSION OF RESULTS 34

Reliability 34

E/I vs Rating Score 34

Criticality Factors 36

Rater Effect 38

VI. SUMMARY AND CONCLUSIONS 39

Summary 39

Conclusions 40

BIBLIOGRAPHY 42

APPENDIX 45

LIST OF TABLES

TABLE		PAGE
I.	Summary of Criticality Factors	20
II.	Frequency of Specimen Occurrence	22
III.	Summary of Scoring	23
IV.	Analysis of Regression of E/I vs Rating Score	26
V.	Analysis of Regression of "E" vs Rating Score	27
VI.	Analysis of Regression of E >1.55/I >1.55 Ratio vs Rating Score	29
VII.	Analysis of Regression of Weighted E'/I' Ratio vs Rating Score	30
VIII.	Analysis of Variance Testing for Rater Effect	32
IX.	Comparison of Raters	33

LIST OF FIGURES

FIGURE		PAGE
1.	Effective/Ineffective Specimen Rating vs Rating Score	17
2.	Frequency Distribution of Rating Score	24

CHAPTER I

INTRODUCTION

I. GENERAL INTRODUCTION

Since the end of World War II, increasing emphasis has been placed on determining the validity of performance evaluation systems and developing evaluation systems with greater reliability. There is little doubt that this emphasis is well directed.[1] The utilization of the results of evaluation systems range from a basis for promotions to justifying a new training program. Without a measure of the validity of a performance evaluation system to temper any decision based on evaluation results, the decision will be questionable.[2] In addition, the cost of evaluation systems varies widely. In order to justify one system over another, the accuracy or validity must be known and considered. For example, the cost of a ranking system would be greater than the cost of a rating scale method. However, the reliability of a ranking system would tend to be greater than the rating scale method.[3]

---

[1]Randolph S. Driver, "Validity and Reliability of Ratings," Personnel, 17:185, February, 1941.

[2]Joseph Tiffin, Industrial Psychology (New York: Prentice Hall, Inc., 1942) p. 253.

[3]C. E. Jurgensen, "Overall Job Success as a Basis for Employee Ratings," Journal of Applied Psychology, 34:333-37.

The development of the classical evaluation systems "involved the transfer of ratings from haphazard, random, and frequently irresponsible judgements of supervisors made perhaps during the heat of a quarrel, to ratings made calmly, deliberately, (and) systematically . . . ."[4] Mee and others group the classical systems used by industrial firms today into three major classifications: ratings or general trait scales, employee comparison systems, and check lists.[5] The development of the many variations within these classifications was prompted by the desire to increase the validity and/or to simplify the systems and reduce their cost.

The development of new evaluation techniques has been attempted in order to solve two problems in the classical systems; those of low reliability and leniency. The forced choice system has been reasonably successful in solving the problem of leniency by raters and the resulting skewed distribution of ratings. The development of the critical incident technique, the major concern of this thesis, is an attempt to remove performance ratings from the subjective and emotional analysis and move to a more highly objective, and thus, reliable system. By the very nature of the system, the tendency toward leniency is suppressed.

---

[4]Tiffin, op. cit., p. 232.

[5]John R. Mee (ed.), Personnel Handbook (New York: The Ronald Press Company, 1952) p. 289.

The critical incident rating technique, as proposed by John C. Flanagan in 1949, was a major contribution to the technology of performance appraisal.[6] Flanagan's theory was that success or failure in performance of an activity is established by a series of actions relevant to the activity. It follows that any appraisal of performance must be based on the actions observed in job performance. Performance appraisal becomes subjective when the appraiser must not only perceive the actions but must make the transition to evaluation of over-all performance. The questions then become, how is the perception and the transition accomplished and is it biased by nonrelevant actions, variations in stimuli magnitude, or time and its effect on the memory apparatus?

Success in quantifying simple stimuli, such as heat and light,[7] pointed to the possibility of quantifying more complex stimuli such as those involved in evaluation of performance. The application of solutions of psychophysics problems in simple stimuli to performance appraisal was proposed by Whitlock.[8] Whitlock has directed an extensive research to prove the theory and to measure the validity of design.

---

[6]John C. Flanagan, "Critical Requirements, A New Approach to Employee Evaluation," Personnel Psychology, 2:419-25, 1949.

[7]S. S. Stevens, "The Psychophysics of Sensory Function," American Scientist, 48:226-53, June, 1960; and "Problems and Methods of Psychophysics," Psychological Bulletin, 55:117-95, July, 1958.

[8]Gerald H. Whitlock, "The Application of Psychophysical Law to Performance Evaluation" (paper presented at Southern Society for Philosophy and Psychology, Atlanta, Georgia, April, 1961).

Recently, at the Y-12 Plant operated by Union Carbide Corporation, Oak Ridge, Tennessee for the U. S. Atomic Energy Commission, the reliability of performance ratings for hourly machining personnel came under close scrutiny and was found to be questionable. Plant management agreed to the establishment of a new performance evaluation system if its worth could be proven in terms of reliability and ease of administration. A modified critical incident technique using "performance specimens"[9] fits these requirements.

II. STATEMENT OF THE PROBLEM

The purpose of this study is to develop a performance specimen check list for hourly machining personnel which is relevant, reliable, and practical to an extent which will allow its use in a given industrial complex.

In addition, certain rater and rating instrument characteristics are investigated from the standpoint of their effects on the variance of performance. Specifically, the problems are to determine:

1. What is the effect of the distance of a performance specimen from the mean on a continuum of performance?
2. What is the effect of rater differences on the variance of estimate?

---

[9] Ibid.

III. IMPORTANCE OF THE PROBLEM

This study is designed to test, in an industrial environment, the knowledge and theories developed from previous work in the application of psychophysics to performance evaluation.

The ability to put into actual practice a specimen check list technique, the results of which can be relevantly quantified, will prove the worth of previous investigations for other than "knowledge for knowledge sake."

In addition, investigation of the factors contributing to the variance of estimate will shed more light on the mechanism of performance evaluation. The importance of this research is self-evident.

IV. DEFINITION OF TERMS

The following definitions will be maintained throughout this study.

Form

Form refers to the performance specimen check list and the category scale as developed for use in rating hourly machining personnel at the Y-12 Plant (see Appendix).

Score

Score is the rated performance value as determined from the category scale in Part II of the form.

Performance Specimen

A performance specimen is "an incident of relevant performance which is uncommonly effective or uncommonly ineffective."[10] A complete set of performance specimens will include all critical incidents.

E and I

For each form the sum of the effective specimens checked by the rater is designated by the letter "E". The letter "I" is the sum of the ineffective specimens checked by the rater.

Criticality Factor

A criticality factor is the measure of the importance of a performance specimen in relation to the total specimen population. It is expressed in distance from the mean on a continuum of performance.

V. APPROACH TO THE PROBLEM

The development of a relevant and reliable critical incident performance rating system must begin with the development of a rating form which incorporates relevant performance specimens. These specimens were collected by interviewing supervisors who are responsible for the judgement of performance. Each supervisor was then asked to rate several subordinates on a rating scale and to check the

---

[10]Gerald H. Whitlock, "The Application of Psychophysical Law to Performance Evaluation," Journal of Applied Psychology, 47:15-23, 1963.

collected performance specimens which they had observed in the performances of the subordinate.

After the forms were scored, the reliability and relevance of the performance specimens in predicting the evaluation of performance was tested. If a high degree of correlation could be found the form would be considered usable.

Criticality factors were determined and used to adjust the performance specimen scores. The effect on the variance of estimate was then examined and if a significant decrease was found, the concept of criticality and differences in specimen weights affecting the transition of observation to evaluation of performance would be strengthened.

Statistical analysis techniques allowed the investigation of rater effects with the data already collected as outlined in the previous paragraphs. If rater differences were detected, the causes would be sought by interviews with the rater's supervision.

VI. ORGANIZATION OF THESIS

The statement of the problem, its importance, the approach to its solution, and the definition of terms have been discussed in this chapter. Chapter II will discuss, briefly, findings of previous studies related to the problem. Chapter III discusses the methodology, Chapter IV outlines the results, and Chapter V is a discussion of the results. The summary and conclusion are discussed in Chapter VI.

CHAPTER II

HISTORY OF THE PROBLEM

This chapter will cover briefly the history of the critical incident performance rating technique and the application of psychophysics to performance appraisal. Wilcox[1] discusses in detail the history of critical incident performance ratings in his thesis entitled The Reliability and Relevance of a Modified Critical Incident Performance Evaluation System. Whitlock[2] has published a detailed summary of the results of the application of the psychophysical law as it applies to performance evaluation. These two references cover the subject and will not be repeated in any detail in the present dissertation.

I. CRITICAL INCIDENT TECHNIQUE AND THE PSYCHOPHYSICAL LAW

The critical incident theory as developed by Flanagan and others during World War II was quickly applied to performance appraisal. The critical incident list developed into critical job requirements which was a great improvement over

---

[1]William J. Wilcox, Jr. The Reliability and Relevance of a Modified Critical Incident Performance Evaluation System, A.E.C. Research and Development Report, K-1392 (Oak Ridge, Tennessee: Union Carbide Corporation, Nuclear Division, 1958), pp. 11-18.

[2]Gerald H. Whitlock, "Application of the Psychophysical Law to Performance Evaluation," Journal of Applied Psychology, 47:15-23, 1963.

the standard list of desirable traits of human beings.[3] In 1956, a critical incident check list was developed at the Oak Ridge Gaseous Diffusion Plant by the Wage Standards Department and in consultation with Whitlock.[4] No attempt was made to quantify the results of this form until Wilcox[5] in 1958 and Beard[6] in 1960 investigated the relation between the items checked and the over-all performance evaluation. Both these studies work with an "effective minus ineffective" score as a linear and logarithmic function, both of which were found to be inappropriate.

In late 1960, Whitlock began investigating the relationship as a power function after reading Stevens' article on psychophysical relations.[7] Since that time Whitlock[8] has directed research which has verified the relationship between observation of performance and evaluation of performance as a power function.

Now that the relationship can be equated and the variance of estimate evaluated, it appears possible that performance appraisal can be predicted from

---

[3]John C. Flanagan, "The Critical Incident Technique," Psychological Bulletin, 51:327, 1954.

[4]Bobby Beard, "Analysis of the Relation Between the Number of Effective Performance Incidents and Ineffective Performance Incidents and the Rated Quality of Performance," (Unpublished Master's Thesis, University of Tennessee, Knoxville, 1960) p. 15.

[5]Wilcox, op. cit., pp. 1-50.

[6]Beard, op. cit., pp. 1-78.

[7]S. S. Stevens, "The Psychophysics of Sensory Function," American Scientist, 48:226-53, June, 1960.

[8]Whitlock, loc. cit.

observation values alone. Indeed, it will be unnecessary to ask for an over-all evaluation of performance except as a continual check on the accuracy of the coefficients of the equation.

II. CRITICALITY FACTORS

In order to define further the relationship which has been verified, Whitlock proposed[9] additional research into the variance of estimate and particularly into the theory that performance specimens have significantly different weights which will effect the evaluation of performance to varying degrees. The only investigation in this area was conducted by Stevenson[10] in 1963. He chose twelve statements from a list of 2000 which had been ranked and weighted by Uhrbrock.[11] The twelve statements, covering the complete range of performance from highly effective to highly ineffective, were ranked by students and scale values calculated using a constant sum ratio estimation.[12] The student scale values were correlated

---

[9] Ibid., p. 23.

[10] H. K. Stevenson, "Factors Influencing the Transition From Observation of Performance to Evaluation of Performance," (Unpublished Master's Thesis, The University of Tennessee, Knoxville, 1963).

[11] Richard S. Uhrbrock, "2000 Scaled Items," Personnel Psychology, 14:375-420, 1961.

[12] Warren S. Torgenson, Theory and Methods of Scaling (New York: John Wiley and Sons, 1958) p. 105.

with the Uhrbrock weights resulting in a linear relationship. However, the function of best fit was not entirely satisfactory.[13]

The twelve statements were also grouped in various combinations and presented to students as observed performance. The students were asked to rank and evaluate the observed performance. Stevenson found a relationship between the additive weights and the evaluation of performance but concluded that it was a result of the experimental design and was not an implication that evaluation is a function of the varying ascribed weights.[14]

---

[13]Stevenson, op. cit., p. 63.

[14]Ibid., p 78.

CHAPTER III

METHODOLOGY

As previously stated, the problem of this thesis is to develop and establish a performance specimen check list in an industrial organization and to study certain factors that contribute to the variance of estimate of evaluations from observation of performance. Methods of approach were as follows:

1. Collect performance specimens and develop a performance specimen check list.
2. Determine "criticality factors" for the performance specimens.
3. Select a sample group of machinists.
4. Complete a performance specimen check list and an over-all rating for each employee in the sample group.
5. Score the form.
6. Analyze data according to:
   a. Reliability and relevance of form
   b. Effect of criticality factor
   c. Rater effects

I. DEVELOP CHECK LIST AND FORM

Interviews were conducted with thirty-six first line production foremen in groups of two to four each. They were asked, "Give me an example of something you have seen an employee do on the job that caused you to say to yourself, 'That man is doing a very good job,' or, 'That man is doing a very poor job'."

The resulting "raw" performance specimens were then sorted and either discarded, accepted, or rephrased, depending on their clarity of meaning and apparent relevance to the job.

The final list of specimens was then incorporated into a three part form. Part I is the list of specimens; Part II is a seven category rating scale divided into 50 points; and Part III is a request for additional performance specimens.

II. CRITICALITY FACTORS

The determination of criticality factors for the performance specimens was approached from two directions. Uhrbrock[1] has developed weights for 2000 performance statements by a ranking technique. The ranking was performed by 120 judges. The selected performance specimens developed for this study were matched as closely as possible with the Uhrbrock statements and the Uhrbrock weights

---

[1]Richard S. Uhrbrock, "2000 Scaled Items," Personnel Psychology, 14:375-420, 1961.

assigned. In some cases the performance specimens were reworded to match the Uhrbrock statement more closely.

The Uhrbrock weights ranged from 10 to 110 with an average of 60. It was assumed that this range represented an almost complete continuum of performance with a spread of $\pm$ 2.50 σ. The standard deviation thus equaled 20. The Uhrbrock weights were then converted to a μ value representing the distances from the mean by $\mu = \frac{wt - 60}{20}$. The performance specimens were then rated by several front line foremen by assigning from one to five points depending upon the importance of the specimen. The average of the foremen's weights were then correlated with the converted Uhrbrock weights and a linear least squares regression developed. The final weights assigned to the performance specimens were then converted to Uhrbrock weights as predicted by the average foremen's weights from the line of regression.

III. SAMPLE SELECTION

The selection of employees to be rated was not a random selection. Employees within the production areas are frequently transferred among areas and work under different foremen. Transfers are made on shop work-load requirements and rarely are based on employee ability. Those employees who had worked under several foremen during the preceding twelve months were used as the sample. A complete list of machinists was sent to each foreman. He was to check all men who had worked under him during the last year and whom he felt qualified to rate.

The sample group of machinists was then chosen in order to provide at least two, and in some cases three, ratings per selected machinist. The sample included the same ratio of machine operators, production machinists, and experimental machinists (the three employee classifications) as was present in the population as a whole.

IV. RATING THE SAMPLE

A total of 280 forms was distributed to 24 foremen with the foreman's and the ratee's name typed on the form. No oral instructions were given. Only the directions on the form itself were used in order to simulate the manner in which the form might later be used. A cover letter was sent to each foreman stating that the ratings were not to go into the employee's personnel folder and that they were not to be reviewed by higher supervision.

V. SCORING

A total of four values was obtained from each form. A rating score value was obtained from the category scale in Part II of the form. The range of the scale was from 1 to 50 with 25 equal to the middle of the average category. This score was to act as a measure of performance and was to be used as a criterion.

An E/I ratio was computed for each form using all performance specimens checked. On forms with E or I values equal to zero, a value of 1 was used to compute the ratio.

An attenuated ratio was computed by summing only E and I specimens checked with predicted Uhrbrock weights or criticality factors greater than 1.55. The 1.55 cut-off value presumably reduced the specimens to 6.1 per cent at each end of the total performance continuum.

A ratio of E'/I' was computed by totaling the individual criticality factors for each E and I specimen checked. On forms with E' or I' equal to zero, one was added in order to compute the ratio.

VI. TECHNIQUES OF ANALYSIS

Linear, least squares regressions were computed and regression analyses were conducted to test the degree of relationship between the various ratios and criterion score. All fits were assumed to be in the form of ln (score value) = ln α + β ln (ratio value). This fit has been substantiated by others[2] and appears to hold true for this study (see Figure 1).

To determine if the residual variance was reduced by the addition of criticality factors, correlation coefficients were compared using a Fisher transformation.[3]

The reliability of the E/I ratio was computed using an independent rater correlation as was the reliability of the foreman weights assigned to the specimens.

---

[2]Whitlock, op. cit., p. 22.

[3]Owen L. Davies, Statistical Methods in Research and Production (New York: Hafner Publishing Company, 1957) p. 193.

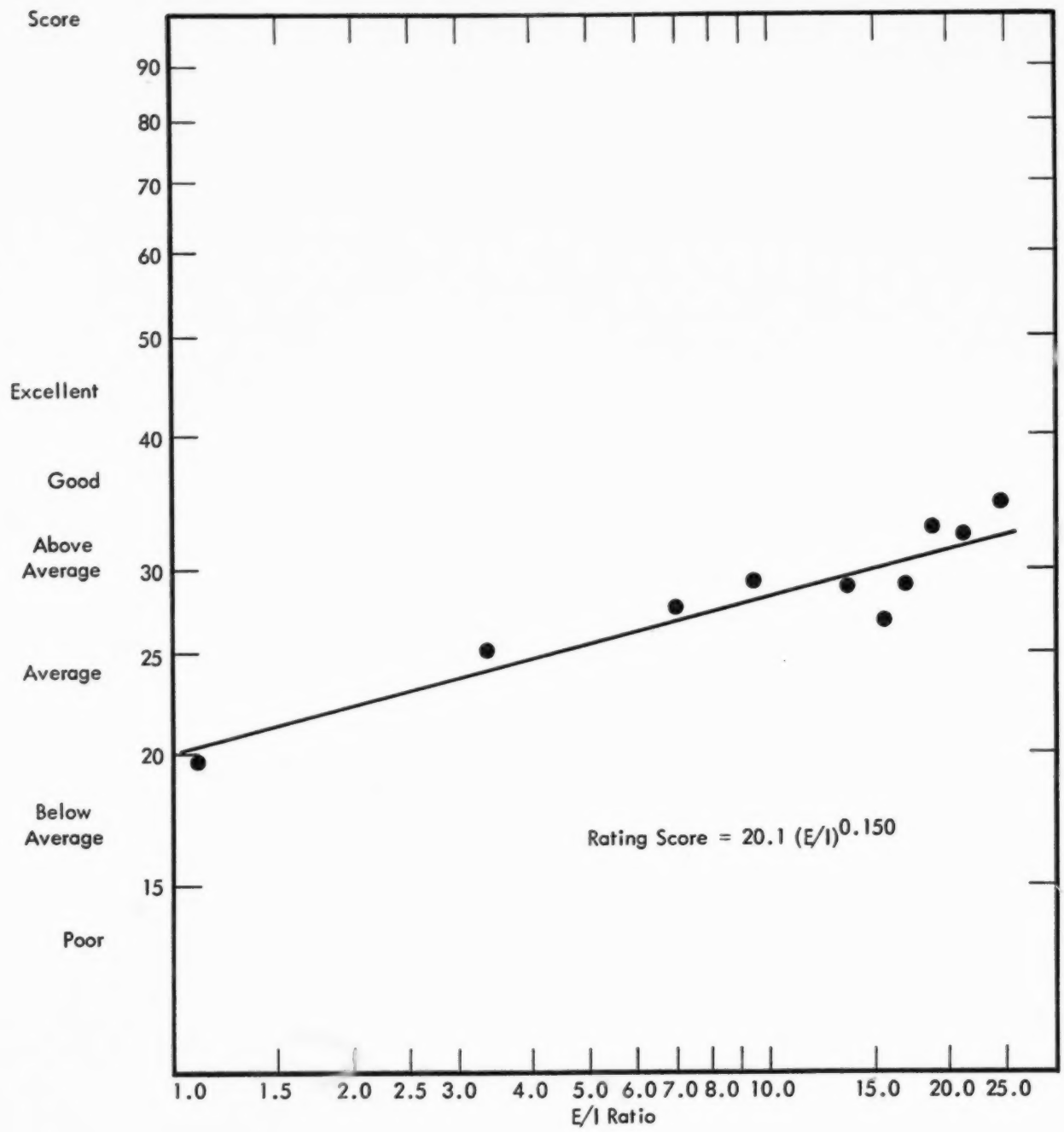

FIGURE 1

EFFECTIVE/INEFFECTIVE SPECIMEN RATING vs RATING SCORE

The correlation between average foreman weights and Uhrbrock values was tested using a Pearsonial correlation.

An analysis of variance was conducted to determine the extent of any rater differences. Since a significant rater effect was detected, a regression analysis was conducted on each rater to determine if a significant regression existed, and if differences in regression coefficients were significant.

Interviews were conducted with shop supervision above the foreman level to determine if the rater effect could be explained by the foreman's ability or attitude differences.

CHAPTER IV

RESULTS

I. THE FORM

The interviews of foremen resulted in over 500 performance statements, the vast majority of which were repeated by many of the foremen. After sorting and rephrasing as clearly as possible, the final list was reduced to 26 effective and 26 ineffective specimens (see Appendix).

A total of 280 forms was distributed. All forms were returned over a four-week period, but only 198 were completed. Those forms which were not completed were returned by foremen who were unable to rate that machinist. In all cases the foreman had earlier replied that the man had worked for him during the last year and that he was capable of judging his performance.

II. CRITICALITY FACTORS

The Uhrbrock weights matched with each of the performance specimens and the average foreman weights assigned by the foremen can be seen in Table I. The reliability coefficient of the foreman weights was found to be only $r = 0.320$ (least significant value at 0.05 level, $r = 0.270$). However, the correlation of the average foreman weights and the Uhrbrock weights provided a coefficient of $r = 0.625$. The predicted Uhrbrock weights as seen in Table I were developed from

TABLE I

SUMMARY OF CRITICALITY FACTORS

	E or I Value	Uhrbrock Weights (μ)	Average Foreman Weight	Predicted Uhrbrock Weight
1	E	1.47	4.0	1.72
2	E	1.72	2.3	1.22
3	I	1.75	3.3	1.51
4	E	1.56	4.3	1.81
5	E	1.81	3.3	1.51
6	E	1.94	3.3	1.51
7	I	1.78	3.7	1.63
8	E	1.56	2.0	1.13
9	E	1.75	3.7	1.63
10	I	2.03	4.7	1.93
11	I	2.00	5.0	2.02
12	E	1.50	2.0	1.13
13	I	2.16	4.7	1.93
14	I	1.56	4.7	1.93
15	E	1.75	4.3	1.81
16	E	1.68	3.3	1.51
17	I	2.03	3.3	1.51
18	E	1.88	3.7	1.63
19	I	1.91	3.7	1.63
20	I	1.12	2.0	1.13
21	E	1.91	5.0	2.02
22	E	1.75	3.7	1.63
23	I	1.69	4.7	1.93
24	I	1.90	5.0	2.02
25	I	1.25	2.7	1.43
26	I	1.81	4.3	1.81
27	E	1.44	2.0	1.13
28	E	1.25	4.7	1.93
29	E	1.75	3.7	1.63
30	I	1.81	4.7	1.93
31	I	1.62	4.0	1.72
32	E	1.56	4.0	1.72
33	I	1.59	4.0	1.72
34	E	2.09	4.7	1.93
35	I	1.22	3.3	1.51
36	E	1.94	4.0	1.72
37	I	1.59	3.3	1.51
38	I	1.56	3.3	1.51
39	I	1.84	5.0	2.02
40	E	1.91	4.0	1.72
41	I	1.72	4.7	1.93
42	E	1.25	3.3	1.51
43	E	1.75	3.7	1.63
44	I	2.06	3.3	1.51
45	I	1.00	4.7	1.93
46	I	1.34	4.7	1.93
47	E	1.66	3.3	1.51
48	E	1.91	4.0	1.72
49	E	1.81	4.7	1.93
50	I	2.47	5.0	2.02
51	E	1.66	4.0	1.72
52	I	1.41	3.7	1.63

the least squares equation $Y = 0.533 + 0.297 X$ where Y is the criticality factor (predicted Uhrbrock weights) and X is the average foreman weights. The average criticality factor for the effective specimens was 1.63 while the average for the ineffective specimens was 1.74. These values are not significantly different. The average criticality factors for the attenuated E and I specimens were 1.76 and 1.87, respectively and are not significantly different. The frequency of observation of the specimens, as determined by the number of checks on the rating forms, may be seen in Table II. The frequency of the E's and I's was correlated with the criticality factors. The ineffective observations had a significant correlation coefficient of $r = -0.450$ while the effective specimens did not have a significant value.

III. SCORING

The results of the three scoring techniques are summarized in Table III. The average rating score for all ratings was 28.6, slightly above the middle of the average category. The distribution of the rating scores is plotted in Figure 2. The standard deviation of the rating score was 6.4.

No new performance specimens were found in answer to Part III of the form. Those replies that were made were very general and vague. A typical reply was "He is a good worker but a little slow."

TABLE II

FREQUENCY OF SPECIMEN OCCURRENCE

Specimen Number	Frequency	Percent of Ratings	Specimen Number	Frequency	Percent of Ratings
1	141	71.2	27	146	73.7
2	98	49.5	28	111	56.0
3	15	7.6	29	73	36.9
4	121	61.1	30	20	10.1
5	158	79.8	31	14	7.1
6	136	68.7	32	144	72.7
7	20	10.1	33	20	10.1
8	101	51.0	34	110	55.5
9	60	30.3	35	15	7.6
10	5	2.5	36	130	65.6
11	12	6.1	37	8	4.0
12	132	66.7	38	13	6.6
13	7	3.5	39	5	2.5
14	0	0.0	40	85	42.9
15	144	72.7	41	18	9.1
16	128	64.6	42	77	38.9
17	11	5.5	43	145	73.2
18	92	46.5	44	18	9.1
19	3	1.5	45	5	2.5
20	8	4.0	46	2	1.0
21	124	62.6	47	57	28.8
22	174	87.8	48	103	52.0
23	6	3.0	49	119	60.1
24	3	1.5	50	2	1.0
25	24	12.1	51	55	27.8
26	2	1.1	52	13	6.6

TABLE III

SUMMARY OF SCORING

Number of Raters	22	
Number of Ratees	95	
Number of Ratings	198	
Average Score (Scale Rating)		28.6
Unweighted Data		
Average Number of E's		14.95
Average Number of I's		1.94
Average E/I		9.04
Attenuated Data		
Average Number of E's		6.95
Average Number of I's		1.24
Average E/I		4.79
Weighted Data		
Average E Value		24.86
Average I Value		2.39
Average E'/I'		12.4

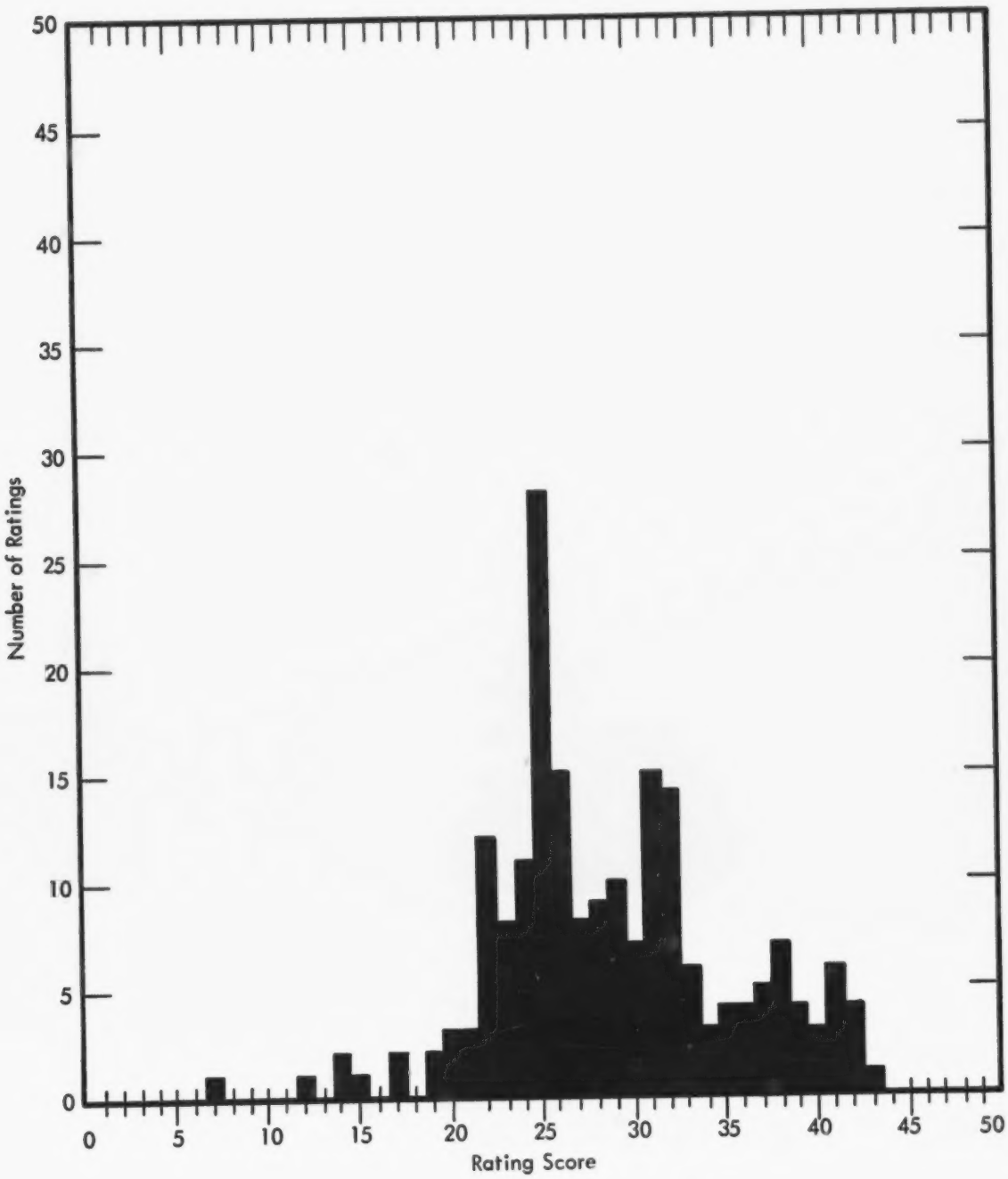

FIGURE 2

FREQUENCY DISTRIBUTION OF RATING SCORE

IV. UNWEIGHTED DATA

The results of the regression analysis for the E/I ratio vs Score may be seen in Table IV. The regression is highly significant with a correlation coefficient of $r = 0.700$. (Note that all regression analysis data are in the form of natural logarithms.) The least squares equation proved to be linear in the form of ln (Score) = ln 20.1 + 0.150 ln (E/I). The goodness of fit can be seen in Figure 1.

When the regression analysis for E vs Rating Score (Table V) was computed, a correlation coefficient of $r = 0.355$ was obtained. When the correlation coefficient for the E/I ratio vs Rating Score was compared with this value, by use of a Fisher transformation, the E/I was found to correlate to a significantly higher degree than just the E. The correlation of I vs Rating Score was found to be $r = -0.627$.

An independent rater correlation revealed a reliability coefficient of $r = 0.663$ for the E/I values.

The correlation coefficient can be corrected for the unreliability of both the E/I predictor and criterion thus giving a relevance estimate.[1] Since no estimate of the reliability of the Score values was available and in order not to overestimate the relevance, the reliability was considered perfect. The correction resulted in a relevance estimate of 0.869.

---

[1]Robert L. Thorndike, Personnel Selection, Test and Measurement Techniques (New York: John Wiley and Sons, Inc., 1949), p. 105.

TABLE IV

ANALYSIS OF REGRESSION OF E/I VERSUS RATING SCORE

$Y = 20.1\ X^{0.150}$
(ln Y = ln 20.1 + 0.150 ln X)

Source of Variance	DF	Sum of Squares	Mean Squares	F Value
Due to Regression	1	6.033	6.033	194.61
Deviation about Regression	196	6.029	0.031	
Total	197	12.062		

Table Value

$F_{0.95}\ (1, 196) = 3.89$

$r = 0.700$

TABLE V

ANALYSIS OF REGRESSION OF "E" VERSUS RATING SCORE

$Y = 21.4\ X^{0.124}$
(ln Y = ln 21.4 + 0.124 ln X)

Source of Variance	DF	Sum of Squares	Mean Squares	F Value
Due to Regression	1	0.452	0.452	16.603
Deviation about Regression	115	3.129	0.027	
Total	116	3.581		

Table Value

$F_{0.95}\ (1, 115) = 3.93$

$r = 0.355$

V. ATTENUATED DATA

When the performance specimens with a criticality factor less than 1.55 were dropped from the form, the effective specimens were reduced from 26 to 17 while the ineffective were reduced from 26 to 18. The average E/I ratio dropped from 9.04 to 7.79, as can be seen in Table III.

The regression analysis (Table VI) yielded a highly significant regression with a correlation coefficient of $r = 0.679$ and a regression line of $\ln Y = 21.0 + 0.184 \ln (E/I)$. The correlation coefficient is not significantly different from that obtained from the total form; however, the slope of 0.184 is significantly greater than 0.150 which was obtained from the unattenuated data.

VI. WEIGHTED DATA

When the ratio of effective to ineffective specimens was computed by summing the criticality factors, the average E'/I' value was 12.4, slightly greater than the ratio for the unweighted data. The data (Table VII) yielded a highly significant regression of $Y = \ln 20.1 + 0.131 \ln (E'/I')$ with a correlation coefficient of $r = 0.693$. As in the case of the attenuated data, no significant difference could be found when comparing the correlation with the unweighted data. The slope of the regression, however, is significantly less than the unweighted at the 97.5 per cent confidence limit.

TABLE VI

ANALYSIS OF REGRESSION OF E > 1.55/I > 1.55
RATIO VERSUS RATING SCORE

$Y = 21.0\ X^{0.184}$
(ln Y = ln 21.0 + 0.184 ln X)

Source of Variance	DF	Sum of Squares	Mean Squares	F Value
Due to Regression	1	5.621	5.621	170.33
Deviation about Regression	196	6.441	0.033	
Total	197	12.062		

Table Value

$F_{0.95}\ (1, 196) = 3.89$

$r = 0.679$

TABLE VII

ANALYSIS OF REGRESSION OF WEIGHTED
E'/I' RATIO VERSUS RATING SCORE

$Y = 20.1\ X^{0.131}$
(ln Y = ln 20.1 + 0.131 ln X)

Source of Variance	DF	Sum of Squares	Mean Squares	F Value
Due to Regression	1	5.824	5.824	183.134
Deviation about Regression	196	6.238	0.032	
Total	197	12.062		

Table Value

$F_{0.95}$ (1, 196) = 3.89

r = 0.693

VII. RATER EFFECT

The analysis of variance to determine if there exists a significant rater effect may be seen in Table VIII. The ΣY^2 is the sum of squares (all data in natural logarithms and standardized about the mean) before adjustment is made for the regression and the $\Sigma Y'^2$ is the sum of squares after adjustment.

When the mean square values for between raters and within raters were compared using an F test, a significant difference was found. Since the between-rater variance is significant, there does exist a difference between the levels of the regression for individual raters. In order to see the magnitude of the difference between raters an analysis of regression was computed for each rater. The results may be seen in Table IX.

Interviews with the raters' supervision revealed little in the way of explaining any of the significant differences between raters. The only relationships found were from statements that raters numbered 11, 12, and 19 have difficulty in differentiating between good and poor performance and that they tend to call all performance "average." As can be seen in Table IX, the data for these raters did not correlate significantly.

TABLE VIII

ANALYSIS OF VARIANCE TESTING FOR RATER EFFECT

	DF	ΣX^2	ΣXY	ΣY^2	DF	$\Sigma Y'^2$	Mean Square
Between Raters	21	55.38	6.02	3.54	21	3.01	0.143
Within Raters	176	213.11	34.23	8.52	175	3.02	0.017
Total	197	268.49	40.25	12.06	196	6.03	

$F = 8.314$

Table Value

$F_{0.95}\ (21, 175) = 1.60$

TABLE IX

COMPARISON OF RATERS

Rater Number	Number of Ratings	Equation	Significant Regression at 0.95	Exponent Different from Zero at 0.975	Exponent Different than 0.150 at 0.975
1	14	$Y = 19.8\ X^{0.094}$	Yes	Yes	Yes
2	3		No(1)	No	
3	12	$Y = 23.0\ X^{0.148}$	Yes	Yes	No
4	8	$Y = 19.9\ X^{0.140}$	Yes	Yes	No
5	6	$Y = 12.0\ X^{0.354}$	Yes	Yes	No
6	11	$Y = 25.7\ X^{0.166}$	Yes	Yes	No
7	8	$Y = 14.0\ X^{0.202}$	Yes	Yes	Yes
8	2		No(2)	No	
9	15	$Y = 19.6\ X^{0.167}$	Yes	Yes	No
10	13	$Y = 21.4\ X^{0.134}$	Yes	Yes	No
11	5		No	No	
12	7		No	No	
13	17	$Y = 19.4\ X^{0.209}$	Yes	Yes	No
14	10	$Y = 22.2\ X^{0.127}$	Yes	Yes	No
15	7	$Y = 13.4\ X^{0.300}$	Yes	Yes	No
16	11	$Y = 15.8\ X^{0.184}$	Yes	Yes	No
17	14	$Y = 18.2\ X^{0.129}$	Yes	Yes	No
18	9	$Y = 14.0\ X^{0.302}$	Yes	Yes	No
19	9		No	No	
20	9	$Y = 13.7\ X^{0.298}$	Yes	Yes	No
21	7	$Y = 15.4\ X^{0.280}$	Yes	Yes	No

(1) Only three points.
(2) Only two points.

CHAPTER V

DISCUSSION OF RESULTS

I. RELIABILITY

The independent rater reliability of the E/I values which was estimated at 0.663 was considerably higher than expected considering the poor reliability[1] of the current performance ratings at the Y-12 Plant. It compares favorably with other inter-rater reliabilities[2] and is significantly above the 0.55 value that Tiffen[3] considers the lower limit for practical use. The time span between the rater's opportunity to observe performances was as great as a year and must have had some degree of effect on the recall ability and thus on the inter-rater reliability. In addition, the performance was observed by raters at different times and thus real changes may have occurred in the performance of the ratee.

II. E/I vs RATING SCORE

The relationship between the E/I ratio and the Rating Score was found to be a power function which is a linear fit on log-log graph paper. The goodness of fit,

---

[1]Unpublished study, undergraduate class in Industrial Management, 453, under direction of Professor Gerald H. Whitlock, University of Tennessee, Knoxville, March, 1962.

[2]Wilcox, op. cit., p. 30.

[3]Tiffin, op. cit., p. 252.

as measured by the correlation of 0.700, is consistent with previous studies of both simulated and actual ratings of technical and education personnel using a critical incident technique.[4] The relevance coefficient of 0.869 must be considered conservative since it is known that the Score values used as a criterion are not perfectly reliable.

The real surprise of this study was the magnitude of the regression coefficient. The slope of 0.150 for the magnitude of the regression coefficient was below that of previous industrial studies.[5] An explanation of the present rating system will shed some light on the cause of the small spread of the rating score and the resulting low slope. Front-line foremen currently rate machinists on a list of attributes such as neatness, attitude, and judgement by checking a box marked excellent, very good, average, or below average. In practice, any rating other than very good or average requires a written or oral explanation which the foremen attempt to avoid. As a result, only in extreme cases will the rating be other than average. With the use of the form developed in this study and the continuous rating scale in Part II which provides a wider range of performance description, the tendency to average will slowly be suppressed and the slope should increase.

---

[4] Whitlock, op. cit., p. 21.

[5] Ibid., p. 22.

III. CRITICALITY FACTORS

The results of the analysis of the attenuated data are too inconclusive to support or discredit the theory that differences in specimen weights are acting on the transition from observation of performance to evaluation of performance. The E/I ratio was significantly reduced when using only the specimens above a 1.55 criticality factor. Since the average criticality factors of the E's and I's were not different on the total form, or on the attenuated specimens, there must exist a tendency to recall and check the effective specimens with low criticality while not checking the ineffective specimens with low criticality. In addition, the slope of the attenuated regression increased due to the reduced range of the E/I values. Thus the attenuated score is more sensitive to the number of E's checked and to the attenuated ratio than it is to the ratio from the total specimen list. It is highly probable that the "halo" effect[6] is operating between the effective specimens and, that by reducing the number of specimens, the redundancy of observed performance descriptions is reduced.

The analysis of the weighted data revealed little difference when compared with the unweighted data. The correlation coefficients of the specimen ratio vs Rating Score were not different and the difference of regression coefficients is in the fringe of significance. It is felt that the reliability of the criticality measure

---

[6]Tiffin, op. cit., pp. 336-40.

is not great enough to allow it to significantly reduce the variance of estimate. Not only was the reliability of the foreman weights low but the original Uhrbrock values had a high standard deviation about the mean weights assigned to each statement.[7]

The correlation of criticality factors and frequency of specimen occurrence support the preceding conclusion. A performance specimen must be uncommon by definition, and it logically follows that the greater the criticality measure, the less the frequency of occurrence. The correlation of the ineffective factors vs frequency of occurrence was low but significant. No correlation could be found with the effective specimen, indicating that the weights of the E's are not operating as assumed. In addition, the frequency of occurrence of some of the effective specimens was as high as 86.7 per cent. A frequency of this magnitude precludes the statement's use as a discriminating performance specimen since it can hardly be considered "uncommon."

When obtaining the foreman weights for each specimen, several specimens obtained as effective and with low weights were considered as highly ineffective by other foremen and therefore discarded. Stevenson encountered a similar reversal when students weighted a portion of the Uhrbrock statements.[8]

---

[7]Uhrbrock, op. cit., pp. 377-420.

[8]Stevenson, op. cit., p. 63.

It becomes apparent that the criticality factors are inherently unreliable and have not reduced the variance of estimate.

IV. RATER EFFECT

The analysis of variance between raters showed a significant rater difference in the level of the regression of E/I vs rating, thus the raters were increasing the variance of estimate to a significant degree. This supports the view that different raters will evaluate similar observations to different degrees.

Attempts to explain the rater differences failed with one exception. The inability to differentiate performance was foreseen in three cases when the data from these raters failed to correlate. This added more "noise" to the system.

CHAPTER VI

SUMMARY AND CONCLUSIONS

I. SUMMARY

This study was the development of a modified critical incident performance evaluation form for hourly machining personnel and an investigation of factors which contribute to the variance of estimate of evaluation of performance from the observation of performance.

Interviews were conducted with front-line foremen to collect performance specimens which were then incorporated into a rating form along with a rating scale for over-all performance. The form was tested by 22 foremen rating a sample group of 198 machinists. The ratio of effective to ineffective specimens proved to be reliable and a relevant measure of over-all performance. The regression equation of the E/I vs Rating Score was consistent with previous results although the slope was less due to the small range of the criterion.

Weights of the importance of the performance specimens as assigned by foremen were found to be low in reliability but the average assignment correlated with Uhrbrock's scale values. A criticality factor was obtained for each specimen by an Uhrbrock value from the average foreman weights.

The incorporation of the criticality factors in calculating the specimen ratio did not reduce the variance of estimate in predicting evaluation from observation.

By reducing the specimen list to items with high criticality factors, the E/I ratio was significantly reduced. This relationship is attributed to a reduction of redundancy and the "halo" effect operating between a portion of the effective specimens.

Rater differences added significantly to the variance of estimate; however, they could not be explained by attitude or ability differences as measured by the foremen's supervision.

II. CONCLUSIONS

The critical incident performance evaluation form developed in this study is reliable and relevant and may be used in the industrial complex for which it was designed. The discriminative ability of the E/I ratio may be increased by eliminating the apparent redundancy. (A measure of the intercorrelation of the specimens will reveal the redundancies.)

The criticality factors appear to describe the general relationship of specimens on the continuum of performance. However, when used in the determination of the effective-to-ineffective specimen ratio they do not decrease the variance of estimate of evaluated performance. These findings are essentially the same as those conclusions reached by Stevenson.[1]

---

[1]Stevenson, op. cit., pp. 77-8.

The increase in the variance of estimate due to rater differences cannot be explained by rater ability or attitudes. It only dramatizes the need for a general equation which will estimate the average performance evaluation response for a given set of observed performances.

BIBLIOGRAPHY

Beard, Bobby. "Analysis of the Relation Between the Number of Effective Performance Incidents and Ineffective Performance Incidents and the Rated Quality of Performance." Unpublished Master's Thesis, The University of Tennessee, Knoxville, 1960.

Davies, Owen L. Statistical Methods in Research and Production. New York: Hafner Publishing Company, 1957.

Driver, Randolph S. "Validity and Reliability of Ratings." Personnel, 17 (February 1941), 185-91.

Flanagan, John C. "Critical Requirements, A New Approach to Employee Evaluation." Personnel Psychology, 2 (1949), 419-25.

Jurgensen, C. E. "Overall Job Success as a Basis for Employee Ratings." Journal of Applied Psychology, 34 (1950), 333-37.

Mee, John R. (ed.) Personnel Handbook. New York: The Ronald Press Company, 1952.

Stevens, S. S. "Problems and Methods of Psychophysics," Psychological Bulletin, 55 (1958), 177-95.

_____. "The Psychophysics of Sensory Function," American Scientist, 48 (1960), 226-53.

Stevenson, H. K. "Factors Influencing the Transition from Observation of Performance to Evaluation of Performance," Unpublished Master's Thesis, The University of Tennessee, Knoxville, 1963.

Thorndike, Robert L. Personnel Selection, Test and Measurement Techniques. New York: John Wiley and Sons, Inc., 1949.

Tiffin, Joseph. Industrial Psychology. New York: Prentice Hall, Inc., 1942.

Torgenson, Warren S. Theory and Methods of Scaling. New York: John Wiley and Sons, Inc., 1958.

Uhrbrock, Richard S. "2000 Scaled Items," Personnel Psychology, 14 (1961), 375-420.

Wilcox, William J. The Reliability and Relevance of a Modified Critical Incident Performance Evaluation System. A.E.C. Research and Development Report, K-1392, Oak Ridge, Tennessee: Union Carbide Corporation, Nuclear Division, 1958.

Whitlock, Gerald H. "Application of the Psychophysical Law to Performance Evaluation," Journal of Applied Psychology. 47 (1963), 15-23.

APPENDIX

EMPLOYEE_____ BADGE NO._____

APPRAISER_____ BADGE NO._____

DATE_____

PERFORMANCE APPRAISAL

This form is used to make periodic appraisals of the demonstrated performance of hourly production personnel on their present job. The form is to be completed by the employee's immediate supervisor. It contains the following parts:

PART I A SERIES OF PERFORMANCE STATEMENTS, the purpose of which is to provide a basis for appraising the employee's performance.

PART II A PERFORMANCE APPRAISAL SCALE to facilitate the appraisal of performance.

PART III ADDITIONAL PERFORMANCE STATEMENTS that were observed by the appraiser but not listed in part I.

INSTRUCTIONS

- Read each Performance Statement in Part I.

  1. If the performance described was observed in the employee's job performance during the last year, place a check mark (✓) in the blank at the left of the statement.

  2. If the described performance is not recalled almost immediately, leave the space blank.

  3. If the performance was not observed during the period, leave the space blank.

- After you have read all the statements and checked those that apply to the employee, rate the employee in Part II.

  1. The Performance Appraisal Scale is divided into fifty (50) points with seven (7) major categories. Place a check mark (✓) at a point which best defines the employee's performance.

  2. The rating should be your opinion of the employee's over-all performance on current jobs.

- In Part III give a brief description of any unusually good or unusually poor performance that you have observed in the employee which has not been described in Part I.

  1. The performance should be a specific action which you have observed and which affected your judgement of the employee.

PART I

PERFORMANCE STATEMENTS

1.____Has been exceptionally prompt to arrive at his equipment and to begin work.

2.____Other men have asked for his advice with work problems.

3.____His tool box, work area and equipment were cluttered and in disarray.

4.____Made an extra effort to meet a deadline when the odds were against it.

5.____Displayed pride in the work turned out by his group.

6.____Displayed ability to adapt quickly to equipment or work with which he was not familiar.

7.____Has reapeatedly left his equipment on unscheduled breaks for personal business or trivial excuses.

8.____Followed instructions cheerfully even when he disagreed with them.

9.____Completed a complex job while working with only his original instructions.

10.____Was abusive to gages, machines, or tools.

11.____Left running equipment unattended.

12.____Was willing and eager to try new equipment.

13.____Displayed a complete lack of interest in his work.

14.____Refused to follow orders to the point of insubordination.

15.____Handled tools and gages as if they were his own.

16.____Displayed patience with a tedious and difficult job.

17.____Deliberately tried to create the impression that he was working when he was killing time.

18.____Displayed competitive spirit in production.

19.____Would not accept advice.

20.____Became aggravated when given a new or different job.

21.____Was able to start a new job with little instruction because he remembered procedures and tooling requirements from a similar job.

22.____Willingly and quickly asked for advice when he needed it.

23.____Damaged a tool or part due to extreme carelessness.

24.____Did not inform his foreman he had damaged equipment but attempted to repair it himself in order to hide his mistake.

25.____Has frequently needed help with a routine job.

26.____Displayed complete disregard for a safety rule.

27.____Displayed concern over killing a part.

28.____Made a genuine effort to improve in an area suggested by his supervisor.

29.____Has displayed alertness to his job that is well above the average.

30.____Has frequently failed to begin work on time.

31.____Did not inform his foreman when his work was delayed or stopped due to equipment or material failure.

32.____Finished his assignment and requested additional work.

33.____Displayed inability to remember instructions.

34.____His quality of work has been exceptionally good.

35.____Worked himself into a corner by not looking ahead.

36.____Displayed ability to capitalize on a mistake and avoid it in the future.

37.____Frequently cut parts of improper speeds or feeds, or in violation of the machining procedure.

38.____Displayed lack of self-confidence to the point that his work was hampered.

39.____Was careless and haphazard in preparing and setting up for a job.

40.____Spotted possible trouble spots by looking ahead in his procedure.

41.____Disturbed the work routine of others by frequent talks and visits.

42.____Exhibited an excellent attitude with respect to constructive criticism of his own performance.

43.____Repeatedly has taken great care to be sure his part is set up properly before he made a cut.

44.____Demonstrated inability to get along with fellow employees.

45.____Would not seek advice when he needed it.

46.____Made the same mistake repeatedly even after given proper instruction.

47.____Made constructive suggestions to improve a machining procedure or fabrication method.

48.____His quantity of work has been exceptionally good.

49.____Displayed pride in his work and questioned reasons for rework.

50.____Attempted to hide a mistake by falsifying production or quality records.

51.____Made a constructive suggestion to improve a machines performance.

52.____Has repeatedly taken up his foreman's time with trivial problems and complaints.

PART II

PERFORMANCE APPRAISAL SCALE

PART III

ADDITIONAL PERFORMANCE STATEMENTS